



Gamified response training with sugary drinks does not facilitate adherence to a restrictive diet

A recommendation by [Zhang Chen](#) ^{id} based on peer reviews by [Matthias Aulbach](#) ^{id} and [Pieter Van Dessel](#) ^{id} of the STAGE 2 REPORT:

Hugo Najberg, Malika Tapparel, Lucas Spierer (2024) The capacity of response training to help resist the consumption of sugary drinks. OSF, ver. 4, peer-reviewed and recommended by Peer Community in Registered Reports.

https://osf.io/eu7j4?view_only=4934c0215f2943cfb42e019792a30b53

Submitted: 26 June 2024, Recommended: 22 November 2024

Cite this recommendation as:

Chen, Z. (2024) Gamified response training with sugary drinks does not facilitate adherence to a restrictive diet. *Peer Community in Registered Reports*, 100860. [10.24072/pci.rr.100860](https://doi.org/10.24072/pci.rr.100860)

Published: 22 November 2024

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Many people nowadays struggle with maintaining a healthy diet. Despite the intentions to eat healthily, they often over-consume highly palatable yet nutrient-poor foods and drinks, which in turn can lead to many health problems. There is therefore a need to develop tools that can help people resist the (over-)consumption of such foods and drinks.

Previous work has shown that training people to execute certain motor responses toward food items can modify their liking for these items, which may also influence their subsequent consumption behavior. Based on these findings, Najberg et al. (2023) developed a mobile game that combined two food-related response training tasks, namely the go/no-go training (Veling et al., 2017) and the cue-approach training (Schonberg et al., 2014). The experimental group was trained to consistently inhibit their responses toward sugary drinks in the go/no-go training, and consistently respond to water items in the cue-approach training (i.e., 100% consistent mapping). In the control group, the mapping between an item and response requirement was 50%, such that participants executed both go and no-go responses toward sugary drinks and water. Najberg et al. (2023) found that after the training, the experimental group reported more reduction in liking for sugary drinks and more increase in liking for water items compared to the control group. However, both groups showed equivalent reduction in self-reported consumption of sugary drinks.

Using the same design (i.e., 100% vs. 50% consistency), in the current study, Najberg et al. (2024) further examined whether the combined go/no-go and cue-approach training game could help people resist the consumption of sugary drinks. Participants were divided into the experimental and control group (N = 100 and

92, respectively), and received the respective training for a minimum of seven days (and up to 20 days). After completing the training, they were asked to avoid the trained sugary drinks. The number of days in which they reported to successfully adhere to this restrictive diet was used as the main dependent variable. Contrary to their predictions, the two groups did not differ in how long they resisted the consumption of sugary drinks after training. Both groups showed equivalent reductions in liking for sugary drinks (contrary to the finding in Najberg et al., 2023), but this reduction in liking was not correlated with the number of successful days of diet in the experimental group. Lastly, those who trained for more days in the experimental group also adhered to the diet for a longer duration, but this correlation might be explained by differences in motivation across individuals.

Together, these results suggest that consistently withholding responses toward sugary drinks and responding to water items does not help people resist the consumption of sugary drinks, compared to a control intervention in which the mapping is 50%. More research is therefore needed to test the effectiveness of food-related response training in changing consumption behavior outside of laboratory contexts.

The Stage 2 manuscript was evaluated over three rounds of review by two expert reviewers who also assessed the Stage 1 manuscript. Following detailed responses to the recommender and the reviewers' comments, the recommender judged that the manuscript met the Stage 2 criteria and awarded a positive recommendation.

URL to the preregistered Stage 1 protocol: <https://osf.io/97aez>

Level of bias control achieved: Level 6. *No part of the data or evidence that was used to answer the research question was generated until after IPA.* **List of eligible PCI-RR-friendly journals:**

- [Advances in Cognitive Psychology](#)
- [Collabra: Psychology](#)
- [F1000Research](#)
- [Journal of Cognition](#)
- [Peer Community Journal](#)
- [PeerJ](#)
- [Royal Society Open Science](#)
- [Studia Psychologica](#)
- [Swiss Psychology Open](#)

References:

1. Najberg, H., Mouthon, M., Coppin, G., & Spierer, L. (2023). Reduction in sugar drink valuation and consumption with gamified executive control training. *Scientific Reports*, 13, 10659. <https://doi.org/10.1038/s41598-023-36859-x>
2. Veling, H., Lawrence, N. S., Chen, Z., van Koningsbruggen, G. M., & Holland, R. W. (2017). What Is Trained During Food Go/No-Go Training? A Review Focusing on Mechanisms and a Research Agenda. *Current Addiction Reports*, 4, 35–41. <https://doi.org/10.1007/s40429-017-0131-5>
3. Schonberg, T., Bakkour, A., Hover, A. M., Mumford, J. A., Nagar, L., Perez, J., & Poldrack, R. A. (2014). Changing value through cued approach: An automatic mechanism of behavior change. *Nature Neuroscience*, 17, 625–630. <https://doi.org/10.1038/nn.3673>

4. Najberg, H., Tapparel, M., & Spierer, L. (2024). The capacity of response training to help resist the consumption of sugary drinks [Stage 2]. Acceptance of Version 4 by Peer Community in Registered Reports. https://osf.io/eu7j4?view_only=4934c0215f2943cfb42e019792a30b53

Reviews

Evaluation round #3

DOI or URL of the preprint: https://osf.io/jckxr?view_only=4934c0215f2943cfb42e019792a30b53

Version of the preprint: 3

Authors' reply, 30 October 2024

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Zhang Chen](#) , posted 28 October 2024, validated 28 October 2024

Revisions Required

Dear Dr. Hugo Najberg,

Thank you again for submitting your revised Stage 2 Registered Report to PCI RR. I think one remaining major issue now is what the *specific* research question of the current study is, and what *specific* conclusion we can draw based on the data.

I appreciate the clarification on the difference between the main research question and the primary hypothesis. However, I agree with reviewer Matthias Aulbach that the research question "Can food response training modify real-world consumption behavior?" is too vague, general, and broad, and therefore needs to be made more concrete and testable in any specific studies.

It is unfortunate that such a vague research question was adopted at Stage 1, and it seems that we now have a disagreement on what the *specific* research question of the current study is. Our (the two reviewers and me) interpretation is that the specific research question is to test the difference between the 100% and 50% contingency groups. I hope you will agree that this is indeed the specific research question raised in the first place. For instance, in the introduction, you wrote (page 4) that "The effect of the intervention was contrasted with a mechanistic control group only differing in the active 'ingredient' of the training: the cue-response mapping rules will be 100% in the experimental and 50% in the control group. This contrast allowed us to control for the confounding factors developed by food cue exposure and cognitive training." For this specific research question, there is a clear answer based on the data: there is no difference between the two groups.

In the response letter, you wrote that "our approach was based on the assumption that the 50% condition would have no effect. This would have allowed us to isolate the absolute effect of the intervention, and thus could answer the main research question". If I understood it correctly, you are now saying that the research question is actually "Can food response training modify real-world consumption behavior, compared to a control condition that has no effect?" However, this is incompatible with what you wrote above, where you said that the reason for using 50% contingency in the control group was to "control for the confounding factors

developed by food cue exposure and cognitive training". So you were saying that the 50% control group might still show some effects due to "food cue exposure and cognitive training", rather than assuming that there would be no effect, and your design allowed you to control for these confounding effects. Second, if the assumption was indeed that the 50% condition would have no effect, this should have been made clear at Stage 1. Furthermore, there should have been a data analysis plan at Stage 1 to explicitly test this assumption. However, this was not the case. Lastly, if the main research question was to "to isolate the absolute effect of the intervention" compared to a control condition that has no effect, I guess there are more suitable control conditions for this purpose, such as using a training that does not involve food stimuli (as in the new proposed study), or a control group that does not receive any training. The 50% control condition was not used to "to isolate the absolute effect of the intervention", so it seems unfair to criticize this control condition, after observing the results, that it "can induce a non-negligible effect of training into the control condition" and therefore could not answer the main research question.

For all these reasons, I respectfully remain unconvinced that the main research question was actually to isolate the absolute effect of the intervention. The original question, "Can food response training modify real-world consumption behavior?", is rather vague and does not explicitly say anything about the absolute effect of the training. Saying that the main research question was actually about the absolute effect of the intervention, rather than the comparison between 100% and 50% groups, sounds dangerously like changing the research question after the results are known, which is exactly one of the biases registered reports aim to guard against.

To sum the long arguments up, your current reasoning seems to be that (1) the primary hypothesis clearly shows no difference between the 100% and 50% group, (2) however, the main research question was actually about the absolute effect of the intervention, and (3) the null results do not allow us to say anything about the main research question. However, as I have tried to argue above, I am not convinced that the research question was actually about the absolute effect (the wording of the question was rather vague, and it did not say anything about absolute effects). Instead, my advice, following what both reviewers have said in their previous comments, is to say (1) there is no difference between the 100% and 50% groups, and (2) explicitly conclude that the answer to the specific research question here is that the supposed active 'ingredient' of the training does not help resist the consumption of sugary drinks. (3) You may still go on and say in applied settings, it may still be interesting to see whether such a training has any absolute effects at all, for instance by comparing it to another control condition. However, it should be clear that the question on absolute effects is different from the specific research question (i.e., the 100% vs. 50% comparison) raised at Stage 1 here. The main conclusion from this research is still that there is no difference between the 100% and 50% condition. This may seem like subtle differences in how to frame the findings. However, I believe this is crucial, because your current reasoning sounds like changing the research question at Stage 2, which is strictly forbidden.

Reviewer Matthias Aulbach reiterated an important point concerning the relationship between devaluation and successful days of dieting. The equivalent devaluation effect between the experimental and control group is now emphasized strongly in the discussion. However, if devaluation is not related to successful days of dieting, to what extent can the null finding on successful days of dieting be explained by equivalent devaluation in both groups? This issue needs to be addressed more carefully.

Lastly, concerning the proposal to conduct a follow-up study, as both reviewers pointed out, the new proposed study will address a different specific question. Regardless of what results you may get from the follow-up study, this will not change the specific conclusion we can draw based on the current data. As such, at least for now, I do not see why this follow-up study should be added as an incremental registration. The proposed study differs from the current study in so many aspects, including (1) the study design, (2) the population, (3) the type of trained items, (4) the behavioral outcome, and (5) most importantly, the specific question being

addressed. For these reasons, I think it makes more sense to submit the follow-up study as an independent Stage 1 manuscript (if you plan to do this also as a registered report), rather than as an incremental registration added to the current Stage 2 manuscript. Independent from whether you eventually decide to go for a new or an incremental submission, I think the issues with the current Stage 2 paper remain, and will need to be carefully addressed first before considering any potential follow-up studies.

Kind regards,
Zhang Chen

Reviewed by [Matthias Aulbach](#) , 23 October 2024

Again, I think the authors have done a good job at addressing my comments. There are, however, still a few points where I am not entirely convinced.

Regarding the distinction between hypothesis and research question, I see the authors' point and agree that it is a good idea to transparently discuss this. In the light of the authors' proposal to run another study, I would, however, recommend caution against this very generally worded research question: I think with any intervention, we need to ask, "is the intervention effective *compared to what?*" (Think: "Does Aspirin reduce headaches compared to doing nothing/taking a placebo/taking Ibuprofen/drinking a glass of water?"). One strength of the current study was to be very specific about this and the new study will be very specific about it, too. That also means that the proposed new study will not be able to answer the more general question because it is a question without an answer per se. Of course, that does not mean that running that other study is a bad idea as it will provide more data relating to the research question.

I might have missed this, but I think one of my points has not been properly addressed. In my earlier comments I wrote "[...] if we assume that devaluation is the (only) mechanism of action that would drive behavioral differences between groups. However, the analyses on hypothesis 2 revealed that changes in liking did not relate to successful days of dieting, indicating that this is not the mechanism by which training would change behavior. This indicates that processes other than devaluation would be driving behavioral effects in both groups." I invite the authors to discuss this issue – devaluation and behavior did not correlate, so why place so much emphasis on devaluation in the interpretation of (null) effects?

Figure 5 and 6: I apologize for spotting this only now, but I think it would be worthwhile to switch the x- and y-axis. I'm aware that the authors computed correlations (which are "non-directional") but I would argue the implicit assumption is that changes in liking/days of training predict the days of successful dieting. Thus, the "dependent variable" successful days of dieting should be on the y-axis, as is common. Of course, the presented information remains the same, it might just be more intuitive to read.

Reviewed by [Pieter Van Dessel](#) , 24 October 2024

The authors have done a good job in revising their paper, clarifying the distinction between their main research question and the primary hypothesis.

I also support their proposal to conduct an additional study as an incremental follow-up to the current research. This new study represents a logical extension as it is well-suited to examine the effects of food response training on consumption behavior.

That said, it is important to emphasize that the proposed follow-up study will not provide insights into the mechanisms underlying how the training produces its effects. The proposed design, by recruiting participants who have already shown difficulties with diet adherence and focusing on consumption frequency, may indeed show changes in behavior. However, as the authors themselves suggest, this does not necessarily speak to how or why these effects occur. It is entirely possible that any type of task involving target stimuli could lead to changes in consumption behavior, regardless of whether those tasks involve specific contingencies or cognitive

processes. This is not inherently problematic if the authors do not intend to make mechanistic claims in the follow-up. The first study can then be noted as a study that allows more specific conclusions in that respect whereas the second study looks at the overall potential of the training in light of the study 1 results.

Evaluation round #2

DOI or URL of the preprint: https://osf.io/7cepq?view_only=4934c0215f2943cfb42e019792a30b53
Version of the preprint: 2

Authors' reply, 21 October 2024

[Download author's reply](#)
[Download tracked changes file](#)

Decision by [Zhang Chen](#) , posted 08 October 2024, validated 09 October 2024

Revisions Required

Dear Dr. Hugo Najberg,

Thank you for submitting your revised Stage 2 Registered Report to PCI RR. The same two reviewers have now reviewed your revised manuscript. Most of the previous comments have been addressed satisfactorily. However, a few issues remain. I would therefore like to invite you to submit a revised manuscript, to address these remaining comments.

1. One main issue, as both reviewers pointed out (and I agree), concerns what the main research question was and what conclusions we can draw based on the data. In the section "The Choice of the Comparator Group Prevents Interpreting the Primary Results" on page 19, you wrote that "our contrast cannot distinguish if the intervention resulted in an absolute increase in participants' capacity to adhere to a diet". This is true, however, this was not the research question raised at Stage 1. Instead, as mentioned in the Introduction, the question was to test the assumed active "ingredient" of the training (100% versus 50% contingency) by keeping other aspects as close as possible. What you framed as a potential risk of "inducing a non-negligible effect of training into the control condition", I would actually argue is a strength, because the research question was to test the difference between 100% and 50% contingency while keeping the other aspects the same, and this control condition allows you to do exactly that. For me, it is thus inaccurate to say that the primary results cannot be interpreted because of the control condition used. Instead, I suggest making it clear that the primary research question was to test the difference between 100% and 50% contingency, and the data give a clear answer to this question: there is no difference between the two groups. I think this is a valid and informative result in itself.

You may discuss why both groups may lead to the same absolute changes (which may explain the absence of any between-group difference), but it should be clear that the question on absolute changes was not the main research question in the first place.

2. Personally, I find the arguments provided in the section "The Role of the Number of Trained Items on the Effect of Response Training" not very convincing. For instance, it is unclear why repeating unhealthy-NoGo pairings for 150 times in the control condition is sufficient to reach the ceiling, while increasing it further to 300 times in the experimental condition makes no difference. Also, it is unclear why repeating unhealthy-Go pairings for 150 times in the control condition has no effect at all. As noted in the previous comments by

one reviewer (Pieter Van Dessel), it is also unclear what underlying cognitive mechanisms these explanations exactly entail.

The two groups had matched expectations in the current experiment. Do you think this can be a more parsimonious explanation for the current results (i.e., whether participants believe the training to be effective or not, see e.g. 10.1016/j.appet.2022.106041)? Did you also measure expectations in your previous work, and did you see a between-group difference there? Do you see correlations between participants' expectations and the effect of training in the current study?

3. For the exploratory analysis on "Diet success rate at each day" on page 17, effect sizes and the 95% confidence intervals are now reported, which is informative. However, the p value of .046 has been omitted, which I think is problematic. By looking at confidence intervals alone, it is often difficult to judge whether the difference is statistically significant or not. Since you claim that there is a difference in failure rate between these two groups, reporting the p value is necessary. I would also recommend explicitly acknowledging here that the analysis is post hoc, the p value is close to the threshold of .05 (thus, the evidence seems weak), and thus this exploratory result should be treated with great caution.

4. The abstract and the conclusion should focus on the confirmatory results. If you mention the results from the exploratory analysis, I would suggest always adding the caveat that it is post-hoc, the evidence is weak, and it should be treated with great caution. And this should also not distract readers away from the main results. For the issue of "the lack of zero-effect comparator", first of all, I do not think it is an issue (see above). Second, it is not relevant for the primary research question, which is about the difference between the two groups, not about the absolute changes brought about by training. As such, I do not think the issue of the control condition belongs in the abstract and the conclusion.

Kind regards,
Zhang Chen

Reviewed by Matthias Aulbach , 25 September 2024

First off, I think the authors addressed many of the comments very well.

However, I respectfully disagree with the authors regarding their interpretation of the null results between the two conditions and the issue around the choice of control condition, even after their revisions following Pieter van Dessel's comments. In their response to Pieter van Dessel's comment, they write "Our main question was whether a 100% association training (experimental group) led to longer diet maintenance than a 50% association." – this research question has a clear answer: it did not.

The authors give explanations as to why that might be the case and their main answer is that participants in the control condition devalued stimuli as much as those in the treatment condition. They then draw the conclusion that "The Choice of the Comparator Group Prevents Interpreting the Primary Results" (page 19). The way I see it, this assessment is only true if we assume that devaluation is the (only) mechanism of action that would drive behavioral differences between groups. However, the analyses on hypothesis 2 revealed that changes in liking did not relate to successful days of dieting, indicating that this is not the mechanism by which training would change behavior. This indicates that processes other than devaluation would be driving behavioral effects in both groups. We can only speculate as to what those mechanisms are (maybe expectations of training effects? Maybe the food exposure? "demand compliant inferences" as Pieter van Dessel suggested?) and I think the authors do a good job at that. However, these alternative explanations do not change the fact that the answer to the main research question is "no".

It is, of course, true that “our contrast cannot distinguish if the intervention resulted in an absolute increase in participants’ capacity to adhere to a diet” (page 19) but that was not the question asked in the first place – the research question did not refer to changes but to differences between two specific tasks. These differences did not emerge, and the Bayes Factor implies equality between groups.

I think an interesting implication of this study’s results then is that the contingency of the pairing does not matter for devaluation to occur. This begs the question: if we were to run the study with a different control condition (say, waitlist control) as the authors propose, what should the intervention look like? Based on the current results, the contingency between stimuli and reaction does not seem to matter (all else being equal).

Regarding the issues around the choice of control condition, I further refer the authors to (Kakoschke et al., 2018). While that article is about Approach-Avoidance Training, the same logic applies here and has been spelled out by the authors, but it might be good to cite this paper, too.

One another note, I think there was a misunderstanding regarding my comment 11: I did not mean to say that the dieting phase impacted the devaluation but merely the instruction to avoid those and/or the participants’ decision to try to avoid them.

Reviewed by [Pieter Van Dessel](#) , 23 September 2024

The authors have done a commendable job revising their manuscript. I have only a few additional suggestions:

1. In the Abstract, the authors write: “We interpret this result as the effect on diet maintenance reaching ceiling in both groups.” This phrasing does not sound very objective or scientific, as it suggests there is only one correct interpretation. I recommend revising this to: “One possible interpretation of this result is that...”

2. In the same sentence, the authors state that this hypothesis (which I would call an interpretation) “is supported by the finding of equivalent target item devaluation in both groups.” I’m missing the logic here. If you are arguing that the effect reached a ceiling, why then support this with a statement that devaluation was equivalent? I believe the authors are not referring to ceiling effects here (and I did not see evidence for ceiling effects, in the sense that nearly everyone maintained their diet). Instead, I think they aim to interpret the null result as suggesting that training in both groups may be equally effective. This interpretation is indeed (to some extent) supported by the equivalent target item devaluation.

3. The following sentence in the Abstract does not logically follow from the previous one, and it is also slightly unclear: “Food response training may also have not improved restrictive dieting adherence in a resourceful, healthy population, as supported by a difference in dieting adherence found only in participants with early failures (18% failure in the experimental group vs. 28.2% in the control group at first quartile).” I suggest rephrasing this to indicate that an alternative interpretation for the null finding (compared to the idea that both trainings impact responses to the same extent) is that there is a difference between the two types of training only in groups that struggle with diet adherence, but this difference is not observed overall due to the small number of participants with such issues in this study. You could then refer to the initial evidence for this by noting the small difference in dieting adherence when considering only participants with early failures.

4. In the Discussion, the title “The Choice of the Comparator Group Prevents Interpreting the Primary Results” seems inaccurate. The choice was actually well-considered, so it’s unclear why it should be reconsidered simply because no effect was observed. The study showed clear evidence that the difference in contingencies is not enough to induce a difference in overall diet maintenance and explicit ratings. That is a clear and valid result and it should be highlighted as such. However, as with every result, there are several possible interpretations. I suggest first stating this result clearly—it seems valid enough, with moderate evidence (looking at the Bayes factors)—and then discussing possible explanations, such as the possibility that the control training had a strong effect.

5. Further in that discussion, the authors state: “However, it entails the risk of inducing a non-negligible effect of training into the control condition. Our design assumed that the control group would always have a lower effect on devaluation than the experimental group and could thus be used for an unequivocal interpretation of the mechanistic effect of an intervention.” I disagree with these statements. A design does not assume anything, and inducing a non-negligible effect in the control condition is not a “risk.” The control group was designed to control for everything except the contingencies, and this is exactly what happened. Hence, the results are clear and they are valid. However, there is also another question: whether response training produces any effect. This question is also a valid question but it is not the focus of the current study as it would need to be answered with a different design. Hence, one should first explain and discuss the key finding (no effect of the contingency manipulation) and only then note that this of course does not mean that response training had no effect. In fact, one possible explanation is that both groups produced similar effects...

6. On page 20, the authors write: “We explain the smaller Group x Session interaction in the current vs. the 2023 study by...” This could be interpreted as the authors being biased toward supporting only one interpretation when there are several others (such as Type I or Type II errors). It would be better to say: “One possible explanation for the smaller Group x Session interaction in the current vs. the 2023 study is that...”

7. On page 21, the authors state: “We interpret this exploratory result...” Here, I would also suggest presenting this as one possible interpretation and I would also advise more caution. For instance: “Although caution is needed because this is an exploratory result with confidence intervals barely reaching significance, one possible interpretation is that...”

8. I noticed the word “expect” used a couple of times when the authors likely meant “except.”

Evaluation round #1

DOI or URL of the preprint: https://osf.io/7cepq?view_only=4934c0215f2943cfb42e019792a30b53
Version of the preprint: 1

Authors’ reply, 04 September 2024

[Download author’s reply](#)
[Download tracked changes file](#)

Decision by [Zhang Chen](#) , posted 19 August 2024, validated 19 August 2024

Decision on your Stage 2 Registered Report: Revisions Required

Dear Dr. Hugo Najberg,

Thank you for submitting your Stage 2 Registered Report, entitled “Sugary Drinks Devaluation with Response Training Helps Early Diet Adherence Failures”, to PCI RR. Two reviewers who have reviewed the Stage 1 RR previously have now reviewed your manuscript. I too have independently read your paper before consulting their comments. As you will see, the reviewers’ assessment is overall positive, but they have also provided some critical feedback which I believe will further strengthen your manuscript. I would therefore like to invite you to submit a revised manuscript, to address the reviewers’ comments.

As both reviewers pointed out, one major issue with the current manuscript is that the interpretation of the

results gives too much weight to an exploratory result. It is certainly fine to report results of exploratory analyses in a registered report, as long as they are clearly labeled as exploratory. However, the main conclusions should then still be based on the pre-registered, confirmatory analyses. Right now, both the title and the abstract focus very much on the exploratory result, which is problematic. These parts need to be revised to accurately reflect the main conclusions from the pre-registered analyses.

Related, the main conclusion on page 18 is that "The current registered report cannot conclude on whether seven to twenty days of combined practice of a Go/NoGo and cue-approach 100% mapping training improve restrictive dieting maintenance in healthy participants when compared to a control group with 50% mapping." However, the statistical evidence for the primary hypothesis (i.e., H1) is clear - there is no significant difference between the experimental and the control group in the number of successful days of diet. Thus, the main conclusion also seems clear - the combined GNG and CAT did not improve restrictive dieting maintenance compared to a 50% mapping control group. Of course, the result might be different if a different control training had been used, or if a different sample had been recruited. However, it is important to note that these explanations are post hoc (i.e., the control training and the sample seemed fine at Stage 1). Furthermore, these are potential explanations for why you did not observe an effect for H1 here. Thus, the main conclusion is still that there is no effect for H1, rather than that no conclusion can be drawn. In general, I think the null findings can be highlighted more in the general discussion (e.g., by linking these to some null findings in the previous literature), and more importantly, the main conclusions should be about the (null) findings from the pre-registered analyses.

The explanations offered in the "The devaluation effect was too large in the control group" section are not entirely clear. Furthermore, alternative explanations exist - as one reviewer pointed out, asking participants to avoid beverages in question may itself change people's evaluation of these beverages. These post hoc explanations need to be clarified, but again, they should not change the main conclusions of the research, which should be based on the confirmatory analyses.

Some of the comments from the reviewers may require revising text that has already been approved at Stage 1 (e.g., the introduction and the methods sections). While I agree that addressing these issues will further increase the clarity of the text, there is also a strict policy on permissible changes between Stage 1 and Stage 2. As such, my advice is to discuss some of the raised issues in the general discussion, but not in the introduction or methods section, such as (1) whether the current measures truly circumvent shortcomings of self-reports, (2) the nature of the correlation between the length of training and the days of successful diet, (3) the reasoning for selecting sugary beverages as the target items etc. The introduction and the methods sections should be kept the same between Stage 1 and Stage 2.

Perhaps contrary to my own advice above, some minor changes to the approved text seem necessary. Note that these changes all concern typographical errors, and are thus permissible changes between Stage 1 and Stage 2. More specifically, some sentences contain grammatical errors, such as:

1. Page 2: "The practice of these tasks have has been shown..."
2. Page 12: "Expectations on the study's hypothesis were also be rated..."

Furthermore, on Page 5: "that MIT interventions can facilitate restrictive diets". The acronym MIT is not defined in the text, nor used anywhere else.

Kind regards,
Zhang Chen

The Stage 2 manuscript “Sugary Drinks Devaluation with Response Training Helps Early Diet Adherence Failures” reports a randomized-controlled trial which showed inconclusive results of a combined cue-approach and Go/No-Go intervention on the length of abstinence from sugary drinks consumption. Further, item devaluation seemed unrelated to abstinence length. Abstinence length and amount of time spent on the training showed a small positive correlation. The manuscript is well written overall and makes relevant contributions to the field. My comments, appearing according to the order in the manuscript, are below.

The timing of the task is spelled out right away for the Cue-Approach Task but not for the Go/No-Go Task. Is this asymmetry intended?

Page 3: “However, whether and how response training intervention impacts consumption behaviors remains largely unresolved.” The authors here introduce the question of “how” effects come about but the rest of the paragraph is not concerned with mechanisms but rather with the “whether” and measurement issues. Regarding the measurement issue, I think this manuscript present only a slight improvement, as it also uses self-report. I understand that soft drink consumption and yes/no questions are probably more reliable to measure than, say, amounts of different kinds of food but some issues of self-report are still not resolved (such as social desirability).

Page 3: “letting the participant stop their training whenever they want in a two-weeks window enables to investigate the link of the intervention’s length on its real-world effect size.” – while this statement seems to avoid implications of causality, I think it still conveys that sense (the preposition “on” carries quite some weight here). It is important to be very clear that self-selected intervention length cannot be interpreted as a causal effect on behavior (as the authors clearly specify in the discussion).

The paragraph on why the authors chose SBBs as a target could be a bit clearer, e.g., portion size is usually unambiguous because of typical packaging and consuming the whole packaging by oneself in one sitting (unlike most snack foods).

Page 5: “Indeed, an additional 5 days of diet (extracted from a Cohen’s d of 0.5 with an estimated standard-deviation of 10 days) would be associated with physiological and cognitive modifications that might be detectable and considered relevant by the participants and the health care providers (i.e., reduction in appetite, higher energy level stability, induction of consumption habits, and realization by the participant that restriction can be maintained).” – could the authors provide a reference for this? As such, it is great that the authors make a substantial argument for a relevant effect size.

Page 5: “Unhealthy participants include self-report of past or current eating disorders, any visual or hearing disability preventing gamified training, and any olfactory or gustative impairment (including smokers consuming ≥ 10 cigarettes daily).” – do the authors mean “ineligible” here instead of “unhealthy”?

Page 6: “Before and after the training, participants rated in a random sequence their 8 most drunk items as well as the water items, from 0 (‘not at all’) to 100 (‘very much’) according to the question ‘Imagine drinking this, how much do you like it?’.” – what exactly does “before and after the training” mean? After each training session? Or after the training phase?

Figure 1: the 50ms delay described in the text is not depicted in the figure.

Page 9: “At the end of the training phase, participants received a weekly questionnaire asking if they succeeded in not drinking the trained sugary drinks and if not, the exact date of the first consumption.” – what exactly does “at the end of the training phase” mean? Maybe the authors could provide a figure with a timeline (at least in the supplemental materials).

The analysis on stimulus liking/devaluation between groups is not presented in the results section but then reported in the discussion. Please also present this in the results section. Regarding this section, it could also be that the mere act of trying to avoid the drinks in question led to devaluation and that this effect is much larger than the training effect and thus hides any training effect.

Page 17/18: again, the authors refer to analyses that were not presented in the results section. Please add this analysis as exploratory to the results section and then refer to it in the discussion.

The interpretation of this analysis seems quite speculative, especially given the arbitrary, post-hoc threshold of 12 days. This should be very clear in the discussion and I don't think it's a good idea to include this in the title (after all, this is a registered report so the manuscript should strongly focus on pre-registered analyses). Those issues set aside, thinking along the authors' line of reasoning, I would argue that response training mainly makes sense in the early phases of behavior change but then we could basically drop it?

I am missing a limitations section. In my view, one important limitation is the focus on participants were willing to abstain completely – that's quite a different sample than those who might be willing to reduce consumption.

Reviewed by [Pieter Van Dessel](#) , 19 July 2024

The authors did an excellent job completing the preregistered study. The manuscript now reports valuable results from a well-designed study.

However, the main limitation of the current manuscript lies in the interpretation.

First, consider the title: "Sugary Drinks Devaluation with Response Training Helps Early Diet Adherence Failures." The procedure does not objectively involve "sugary drinks devaluation." Instead, it examines the effects of completing combined gamified GNG and CAT tasks. Furthermore, it is not accurate to state that completing these tasks "helps early diet adherence failures." This was not the research question examined. The main question was whether completing combined gamified GNG and CAT tasks increases the number of successful diet days. The answer is that it did not. A title that highlights this main result would be more appropriate. The current title focuses on an exploratory result, which is not ideal for several reasons (also see below).

The abstract also gives way too much attention to the exploratory result: "Finally, exploratory analyses indicated that the experimental group had improved diet adherence on participants failing the diet early (18% failure in the experimental group vs. 28.2% in the control group at first quartile). Our collective data seems to indicate that the effects of food response training may be particularly beneficial for individuals with difficulties adhering to diets. We suggest conducting a similar study to validate this exploratory result with a more fitting design." I would suggest removing this and providing a conclusion about the main study results.

Also the discussion has limitations related to the interpretation of results. The title of one paragraph is "The devaluation effect was too large in the control group." It is unclear why the authors chose this title. Objectively, there was a reduction in evaluation in both groups, and this was not significantly different between groups. But this result doesn't imply that the devaluation effect was "too large." Of course control training can also have effects (many studies suggest this), but that is not a problem, in fact it is often a positive outcome.

The authors compare the results to their prior study and note: "The lower number of trained items resulted in a smaller Group x Session interaction." This may not be the case, as there could be other explanations, such as sampling differences. They also state: "the number of NoGo associations in our present control group likely led the associative learning to reach its ceiling." It is unclear why this is "likely" the case. Additionally, the meaning of "associative learning" here is ambiguous. Do the authors refer to a specific cognitive process underlying GNG effects, such as the formation of associations? Clarity is needed.

Furthermore, they note: "the equivalent devaluation between the experimental and control groups in the present study suggests that the effect of unhealthy Go associations did not fully neutralize the effect of unhealthy-NoGo associations." Again, this appears to refer to an associative cognitive process (if I understand well what is meant with "neutralization"). Note however that associative explanations of GNG effects are not well-supported anymore. If the authors want to refer to such explanation, it should at the very least be clarified that there are also other explanations, such as inferential explanations (related to demand compliant inferences, but also other types of inferences). The authors seem to allude to inferential explanation in the next sentence "Indeed the unhealthy-Go associations could have counteracted the unhealthy-NoGo associations if the participants were not expecting the intervention to be effective." Here they refer to expectations (i.e., causal inferences). However, this point is unclear as the potential explanations are not well defined.

The following sentences are also unclear: “Overall, we conclude that our control group did not allow an unequivocal interpretation of the mechanistic effect of an intervention with more than 150 unhealthy-NoGo associations per item. Since we could not be sure that the control group experienced a meaningfully lower effect of training than the experimental group, and without pre-training measures of diet capacity, we cannot conclude on the effect of response training on diet adherence (i.e., the primary hypothesis).” Are the authors suggesting that control training can also have an effect? This is true and has been evidenced by other studies. However, it is unclear why this is relevant. The authors initially posited that the contingency difference is the crucial working mechanism. If they are revising this idea, it should be discussed. This can be done with reference to inferential theories that indicate that learned propositions rather than contingencies that form associations are crucial (and, for instance, provide evidence in reference to instruction-based effects showing that contingencies are not crucial for effects to arise).

The paragraph on “The Role of Participant Baseline Capacity on Diet Adherence” explains one possible reason for the lack of effects for H1: the effect was only present for early maintenance (although it remains unclear why this effect would not be observed overall). However, this explanation does not warrant a separate paragraph. It should be discussed as one possible explanation and nothing more. Currently, this result is given too much attention (see also the abstract and conclusion) despite a p-value that is not robust ($p=.046$) and it being a pattern observed post-hoc and only at a specific moment in time (before 12 days), which resembles data dredging or p-hacking. The authors should be cautious here.

Minor issues:

- Overall, it would be beneficial if the discussion were to explain the three hypotheses and possible explanations in more detail (e.g., motivation is mentioned for H3 but not elaborated).
- Adding Bayes Factors for the null results could provide important information (e.g., is there strong evidence for the absence of an effect?).
- It could be useful to report whether more days of training correlate with the number of successful diet days in the control group and if this correlation differs significantly from the experimental group.
- Several studies in CBM research have found that CBM does not affect real-life behavior in general groups (in contrast to clinical groups; see papers by Reinout Wiers). Discussing this could be beneficial.
- There is not much information about study limitations related to the specific sample, self-report measurement,...